REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
16-12-2011	Final Report	1-Jul-2008 - 30-Jun-2011

4. TITLE AND SUBTITLE

Highly extensible programmed biosensing curcuits with fast memory

5a. CONTRACT NUMBER
W911NF-08-1-0296

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
James Broach, Alexandre Morozov, Ron Weiss

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Princeton University
Office Of Research & Project Administration
The Trustees of Princeton University
Princeton, NJ 08544 -0036

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
54399-LS.2

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The overall aim of the project is to develop a robust platform for an array based detector that could sense, distinguish and quantify diverse collections of environmental analytes. We have previously developed cell based reporters that afford the ability to recognize a large number of chemicals, built around G-protein coupled receptors (GPCRs), which provide high diversity and broad specificity. To render this detector system able to function in real time, we are applying synthetic biology approaches to engineer cells with a fast, phosphorylation based

15. SUBJECT TERMS
biosensors, synthetic biology, Saccharomyces cerevisiae, G-protein coupled receptors, synthetic nose

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			James Broach
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 609-258-5981

Report Title

Highly extensible programmed biosensing curcuits with fast memory

ABSTRACT

The overall aim of the project is to develop a robust platform for an array based detector that could sense, distinguish and quantify diverse collections of environmental analytes. We have previously developed cell based reporters that afford the ability to recognize a large number of chemicals, built around G-protein coupled receptors (GPCRs), which provide high diversity and broad specificity. To render this detector system able to function in real time, we are applying synthetic biology approaches to engineer cells with a fast, phosphorylation based memory circuit. This solves two problems: the readout is based on protein phosphorylation and thus occurs within seconds. Second, the response, once established, remains fixed, so that the readout can be analyzed without a transient loss of signal. In order to interpret the results we obtain from the proposed array detector, we have developed a Bayesian-based computational method for extracting the identities and amounts of compounds in a mixture. Applying our computation approach to results obtained with a prototype GPCR-based array, we were able to extract the identity and amounts of compounds in complex mixtures. This provides validation of the method, which could be of broad use for any array based detector system.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received	Paper
2011/12/02 1 1	Julia Tsitron, Addison D. Ault, James R. Broach, Alexandre V. Morozov. Quantitative decoding of complex chemical mixtures with sensor arrays, PLoS Computational Biology, (10 2011): 0. doi:

TOTAL: 1

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received	Paper

TOTAL:

Number of Manuscripts:

Books

Received	Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Deepak Mishra	0.50	
FTE Equivalent:	**0.50**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Addison D. Ault	0.50
FTE Equivalent:	**0.50**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Ron Weiss	0.10	
James R. Broach	0.10	
FTE Equivalent:	**0.20**	
Total Number:	**2**	

Names of Under Graduate students supported

NAME	PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Foreword.

The overarching aim of the project has been to develop a biosensor array based on the principles of mammalian olfaction. Using the tools of synthetic biology, we worked to create living cells that would serve as sensor elements in such an array and that would possess a fast, phosphorylation based memory circuit, responsive to G protein-coupled receptors (GPCRs) and histidine kinases (HKs) inputs (Fig.1). The phosphorylation based signal response allows a very fast biological readout, so that the biosensor can function in real time, unlike those based on transcriptional readouts. The toggle switch design incorporated into the circuit allows cells to maintain a memory of analyte exposure, which enhances the sensitivity of the sensor to analyte concentration. The use of GPCRs as the analyte receptors in the sensor elements allows enormous versatility in our ability to tune the array to any of a myriad different analytes. Moreover, by using different receptors with overlapping analyte specificities, we can detect a significantly greater number of analytes than the number of distinct sensor elements and can distinguish and quantify individual components presented in complex mixtures. Thus, the format matches the needs presented by real world conditions.

A second facet of the program has been a computational study to develop software that would decode the output of the biosensor. As in the olfactory system, each of the elements of the biosensor array responds to a distinct set of multiple analytes that overlaps the sets recognized by other elements. Accordingly, decoding what analytes were present in an applied mixture that produced a particular pattern of element responses becomes substantially non-trivial as the number of analytes in a mixture increases. Accordingly, a biosensor array is useful only if the pattern observed can be interpreted to yield the identities of the components in the mixture applied to the array. The Bayesian computational method we developed provides identification and quantification of all the components of a mixture tested in an array format and can be generalized to any olfactory-like biosensor array. This computational package also guides the design of any array to optimize the discriminatory capability and to minimize the array components.

As described in this final report, we made excellent progress on the first aspect of the project and completed the second task. Moreover, we have substantially streamlined the synthetic biology process by expanding the toolkit with which we create de novo designed strains. These results will facilitate further applications of synthetic biology to creating specific biology based circuits and sensors and have brought us to the point of reducing the biosensor to practice.

List of Figures and Tables (Provided as a pdf attachment).

Problem Addressed

This project focused on development of a format for cell based biosensors, addressing three critical current shortcomings: 1) engineering the back end of the biosensor by creating a rapid readout of sensor activation; 2) engineering the front end of the biosensor by identification and implementation of suitable receptor elements that would provide broad spectrum coverage of the chemical space of interest; and 3) designing the sensor “brain” that would interpret sensor output to reveal the identities of the sensor inputs and quantify their amounts.

The first part of this program addressed a major problem in designing cell based biosensors, namely, to design a cell based system with a rapid readout. All previously described cell based reporter systems have used a transcriptional readout, which provides colorimetric, fluorimetric or growth readouts. These have proved useful in engineering and optimizing various biological circuits using the tools of synthetic biology and as a format for cell based assays for various drug screening purposes in the pharmaceutical industry. However, transcriptional based readouts are inherently slow, due to the multiple biological

steps required for producing the final reporter product. Thus, in order to create a cell based assay that would provide useful feedback in real time, one needs a new platform for cell response that would transmit information on the presence of a stimulus and provide a detectable output in a very short time. To solve this problem, we proposed to develop a novel signaling and response network in the yeast Saccharomyces cerevisiae based on protein phosphorylation. The rapid in vivo kinetics of protein phosphorylation and the extensive information on natural phosphorylation networks in cellular signaling suggested that this was a feasible approach to solving this problem.

The second problem we addressed in this project was design of the front end of the sensor to allow broad spectrum coverage of chemical space. We proposed to approach this issue by basing the sensors on the family of G-protein coupled receptors (GPCRs). Currently, more than 4000 different GPCR genes have been identified, with specificity for an equally broad number of different chemical compounds. This diversity often allows selection of an individual receptor from the existing repertoire to fit a particular detection need. Moreover, we have recently shown that this diversity can be artificially increased by application of the tools of protein engineering to evolve a particular receptor to recognize a new chemical entity. Finally, the olfactory class of GPCRs exhibit degenerate and overlapping ligand recognition. This degeneracy is the basis of mammalian olfactory perception, allowing a relatively small number of distinct receptors (200-500) through a combinatorial process to recognize and distinguish a very large number (>100,000) of distinct chemical entities. Accordingly, we proposed to design the front end of our biosensor on the basis of the olfactory principle to allow maximum flexibility in application of the biosensor. This aim thus required that we engineer our yeast cell based system to functionally express a broad spectrum of GPCRs and to couple those receptors to the phosphorylation-based signaling network described above.

Our proposed use of combinatorial sensor arrays to detect a large number of analytes using a relatively small number of receptors raised a final problem that had to be addressed, namely, how to interpret the output of such a sensor array to identify the impinging chemical entities. The complex pattern of receptor responses to even a single analyte, coupled with the nonlinearity of responses to mixtures of analytes, makes quantitative prediction of compound concentrations in a mixture a challenging task. While the output of these cross-specific arrays in response to single compounds can generally be interpreted through pattern recognition algorithms, computational analysis becomes more difficult when the array is presented with a mixture of compounds. Indeed, the non-linear nature of sensor responses to multiple ligands makes it hard to train discriminatory algorithms on a ‘‘typical’’ subset of patterns. The non-linear dependence of sensor output on ligand concentrations is generic in reporter systems and may be compounded by potential binding interference of the two ligands, saturation of the sensor output and, of particular concern, potential antagonistic action of one ligand on another’s activity. As a result, responses to complex mixtures have primarily been used to ‘‘fingerprint’’ specific mixtures rather than identify their constituents quantitatively. Accordingly, to achieve success, we needed to develop a novel and robust method for interpreting the output of the sensor arrays.

Major achievements

Enhancement of the tools of synthetic biology: rapid implementation of “plug and play” modules.

Achieving our goal of developing a cell based sensor with rapid response time required extensive application of the tools of synthetic biology, that is, of creating novel combinations of genes in a living cell that would redirect the normal function of the cell to perform a novel task. Given the large number of manipulations often required to re-engineer a cell to a desired novel specification, we spent some effort in developing methods to facilitate such manipulations.

To ease the time consuming process of large circuit construction in Saccharomyces cerevisiae, we have designed, built, and finalized a DNA assembly system for yeast systems. Our approach harnesses the strengths of yeast homologous recombination, a strategy employed for decades in biological research, and couples it to recent advances in synthetic biology stemming from recombination-based cloning strategies [4] and Gibson DNA assembly [5].

The system operates in two stages and pictorially represented in Figure 2:

1. Establishing a single transcriptional unit of promoter and gene.
2. Assembling multiple transcriptional units together into a backbone

For stage one, we start with a standard library of promoters and genes flanked by Invitrogen Gateway attL and attR sites. This library is fully compatible with other parts libraries, notably work of Susan Lindquist’s group at the Whitehead Institute and the plasmid libraries maintained by Harvard Medical School and Arizona State University. The Gateway ‘LR Reaction’ is performed in order to assemble promoter and gene pairs together into a destination plasmid such that the promoter-gene pair is flanked by defined 40-bp sequences. For example two ‘LR Reactions’ might yield the following plasmids: (Seq1 – Promoter – Gene – Seq2) and (Seq2 – Promoter – Gene – Seq3).

For stage two, we utilize the recently published Gibson Assembly relying on sequence and ligation independent cloning. By

having 40bp homology regions on ends of individual DNA fragments, the Gibson reaction allows these 40bp flanked fragments to be combined to yield a single part. Thus we linearize the corresponding plasmids of stage one (Seq1 – Promoter – Gene – Seq2) and (Seq2 – Promoter – Gene – Seq3) and add it to a reaction containing a linearized vector (Seq1 – Vector – Seq3). Upon reaction completion, we obtain a circular vector containing (Seq1 – Promoter – Gene – Seq2 – Promoter – Gene – Seq3). A pictorial representation of this process appears in Fig 2B.

Under the aegis of this program, we have built the infrastructure necessary to implement this strategy, including a family of 18 promoters, 25 genes, and 8 gibson-compatible backbones. The promoter library spans both constitutive and inducible promoters and the gene library includes fluorescent proteins, exogenous yeast components used in previously published S. cerevisiae papers from the Weiss laboratories. The Gibson-compatible backbones include the centromeric pRS plasmids allowing low copy propagation, the 2 micron plasmids allowing high copy propagation, and a novel site-specific integration plasmid. Figure 3 shows the construction of a family of 4 plasmids each containing a different fluorescent protein.

Creation of a rapid readout for cell based receptor activation.

A major goal of this project was to develop a means of rapidly detecting receptor activation in yeast. We have accomplished that goal by adapting the rapid phosphorylation cascade underlying MAP kinase signaling in yeast to yield a fluorometric response to receptor activation. As shown in Figure 4, three endogenous mitogen-activated protein kinase (MAPK) pathways – the mating pathway, the high osmolarity response pathway (HOG), and the filamentation pathway - coexist and function independently in yeast cells. In particular, the phosphorylation cascade of the HOG pathway results in rapid translocation of the Hog1 transcription factor from the nucleus to the cytoplasm in response to activation of the osmo-responsive receptor Sho1. While one can follow the translocation of Hog1 from the cytoplasm to the nucleus in microfluidic devices (Figure 5), this detection does not lend itself to use as a readout in a multi-element sensor array. Accordingly, we have designed and implemented a split GFP format so that activation of the receptor converts cells from non-fluorescent to fluorescent upon pathway activation.

Upon activation of the HOG pathway, the Hog1 transcriptional activator translocates into the nucleus, where it physically associates with the chromatin bound Hot1 protein. Accordingly, we created a strain in which Hog1 is fused to the N-terminal domain of the fluorescent protein eGFP and Hot1 is fused to the C-terminal domain of eGFP. The logic of this design is that in the absence of stimulation, Hot1 and Hog1 reside in different cellular compartments and, as a consequence, the two halves of GFP cannot associate and the cells are non-fluorescent. Upon stimulation, Hog1 relocates to the nucleus, binds to Hot1, allowing the two halves of GFP to associate and fold into a fluorescent protein, rendering the cells fluorescent. In this manner, pathway activation can be detected and quantified in whole cells or cultures of cells by the acquisition of fluorescence in proportion to the degree of stimulation. As shown in Figure 6, we have been able to accomplish this goal. Within five minutes of stimulation, we observe significant fluorescence of cells. This is more than an order of magnitude more rapid than any transcription based reporter assay described to date. In obtaining this rapid cell response, we have achieved one of the major goals of this project.

Coupling GPCR activation to our rapid response readout.

To functionally couple different GPCRs to our rapid readout, we have exploited the yeast cell's endogenous GPCR signaling pathway. Haploid yeast cells express a single GPCR, which is activated by pheromones produced by cells of the opposite mating type and which upon stimulation activates a MAPK pathway resulting in transcriptional activation through the Fus1 transcriptional activation (Figure 4). By engineering the G-protein that bridges the GPCR with the MAPK signal cascade, we have been able to create yeast strains that can functionally couple a variety of different mammalian GPCRs to the MAPK and thereby generate yeast cells that can detect and report on activation of mammalian receptors by their cognate ligands. This list of mammalian GPCRs that have been successfully coupled to the MAPK activation in yeast is provided in Table 1.

While our previous work provides a means of using a variety of GPCRs as the front end of a biosensor, the readout for activation of these receptors has been transcriptional activation. To convert such strains to a rapid readout response, we have exploited our prior success in redirecting the pheromone signaling response into HOG pathway activation. As noted in Figure 4, the pheromone MAPK pathway shares a number of components with the HOG pathway. We previously showed that these two pathways were insulated from each other by mutual inhibition. Accordingly, by eliminating the mutual inhibition, we have been able to channel activation of the GPCR pathway into a HOG pathway response. Since this cross inhibition is mediate by the terminal MAP kinases of the pheromone pathway, Fus3 and Kss1, we have created variants of our reporter strains that lack the genes for both these kinases. In such strains, output of all three MAPK pathways is redirected solely to the high osmolarity pathway (Fig.2). Both, stress response and GPCR activation result in Hog1 phosphorylation and its translocation to the nucleus. This aspect of the program is still undergoing optimization.

Engineering a phosphorylation based toggle switch in yeast.

As a refinement of our cell based detector assay, we have focused on implementation of a toggle switch in the response pathway. By incorporating a toggle switch into our array design, we can generate a biosensor that possesses intrinsic memory. That is, the readout of the cell once exposed to an activating ligand will remain on even after the ligand is removed. This feature allows both assessment of prior exposure to a chemical entity as well as a means of resetting the detector at will. This toggling is achieved by negative coupling of two response pathways, each possessing a positive feedback loop. Accordingly, once one pathway is activated, it both remains activated and inhibits the activity of the second pathway. Stimulation of the second pathway turns off the first and, once this second pathway is activated, it remains activated.

We have proposed implementing a toggle switch by negatively coupling the pheromone responsive MAPK pathway to a Hog1p-responsive JAK2-STAT5 signaling module capable of feeding back to the phosphorelay input of the HOG1 pathway (Figure 1). Our toggle switch involves Jak2 inactivation with phosphorylation of Hog1 and STAT5 activation with Ypd1 phosphorylation. Central to this design are two chimeric proteins: Jak2-Hot1 and Stat5-HKRR. Active Hog1 fused to a phosphatase domain of SHP1, PTP, contacts and inactivates Jak2 as a part of the split-GFP reporter system. When Hog1 is not phosphorylated, Jak2 remains active and signals to Stat5. To enable phosphorylated JAK2-STAT5 to control Ypd1/Ssk1 phosphorylation we designed a Stat5-HKRR fusion protein where cytoplasmic Sln1 is fused to the C-terminal of Stat5. Sln1 is a 1220 amino acid protein with four distinct regions: (a) an N-terminal extra-cellular domain (ECD), (b) a cytoplasmic linker region, (c) a histidine kinase (HK), and (d) an aspartate response regulator (RR) domain [3]. We assume that activation of Jak2 dimerizes two Stat5-HKRR histidine kinase domains. Therefore, activation of JAK2-STAT5 will enable autophosphorylation of two HKs, resulting in suppression of HOG1 pathway through phosphorylated Ypd1 and Ssk1.

To achieve the toggle network topology, the JAK2-STAT5 signaling module must be made responsive to Hog1p. This link is made by producing the chimeric proteins PTP2-Hog1, Jak2-Hot1 and Stat5-HKRR. Sensitivity analysis of a quantitative model of the toggle network has shown that the activation of Ypd1 via Jak2 and Stat5 is critical to bistability. Our initial results showed that Jak2/Stat5 was capable of a 2-fold steady state increase of phosphorylated YPD1 mediated transcription. However, if this steady state fold difference can be increased, the regime of bistability increases dramatically.

The Jak2 protein is comprised of two domains, JH1 and JH2. JH1 is the catalytic domain while JH2 functions as a regulatory domain, inhibiting JH1's function [7]. Thus, we have created a JH1/Stat5 system where instead of utilizing Jak2, JH1 is substituted. Both Jak1/JH1 and Stat5 are under inducible control. Figure 7 shows steady state population level mean fluorescence values. By utilizing JH1, we are able to get a 5.3 fold difference in the steady-state pathway activation. Further experiments are underway to modulate the JH1 and Stat5-HKRR levels to increase apparent gain for use in the toggle network.

Computational Methods for Array Data Deconstruction.

We have addressed a significant and unresolved problem in the development of array based detectors. Many array-based detector formats have been described but all are based on assembling a collection of individual sensor elements, each with non-identical but overlapping sensitivities to the collection of compounds or analytes that the experimenter wishes to detect and quantify. The fundamental problem is how to extract from the pattern of responses of the individual sensors on the array the nature of the stimulant applied to that array. The standard approach to this problem has been to invoke the ability to recognize patterns of activation. That is, if a test compound elicits a particular pattern of activation of a subset of sensors then one concludes that the test compound is equivalent to a reference compound that elicited the same pattern of activation. While this might be adequate for determining single test solutions containing a single compound, this approach is completely inadequate for determining what is present in a mixture of two or more compounds. This is even more problematic if the activities of the compounds on a sensor are not strictly additive, as would be the case for receptor antagonists. As far as we can determine, no one has developed a method for extracting from array data the composition of mixtures of compounds. This is unfortunate, since most of the real-life situations in which such arrays would be used would involve mixtures of compounds.

We have developed a Bayesian-based computational method for extracting the identities and amounts of compounds in a mixture using array based detectors. We previously used directed evolution to generate a collection of variants of the human UDP-glucose receptor, each of which possesses over-lapping but not identical binding and response activities against four separate ligands, UDP-glucose, UDP-galactose, UDP and UDP-glucosamine [5]. Using the yeast-based, transcriptional reporter for measuring GPCR activation, we first calibrated each of the receptors by determining dose response curves for each receptor against each of the ligands. We then applied nested sampling and Bayesian inference to the data to extract a binding affinity and an efficacy value for each of the ligands on each of the receptors. These receptors along with the calibration data constitute the detector array that we used for determining the nature of unknown mixtures.

In the second stage, we prepared various mixtures of the four ligands and applied them to the array. To determine the composition of a mixture, known dilutions of the sample were applied to each of the receptors and the resulting activation determined across the dilution series. These data are then processed by our Bayesian algorithm though nested sampling, which returns the most likely combinations and concentrations of ligands that would yield the observed response curves. This approach was remarkably effective in identifying the compounds present and their concentrations in a variety of different

mixtures. We were accurate to within 20% in determining the presence and concentrations of any mixture of the four ligands. This is a spectacular and unprecedented result that will be broadly applicable to any array based detector system.

References

1. McClean, M. N., Mody, A., Broach, J., and Ramanathan, S., Decision Making in MAP Kinase pathways. Nat. Genet., 2007, 39, 409-414.
2. Magliery, T. J. et al., Detecting Protein−Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism. J. Am. Chem. Soc., 2005, 127 (1), 146–157.
3. Hohmann, S., Osmotic stress signaling and osmoadaptation in yeasts. Microbiol Mol Biol Rev, 2002, 66(2):300–72.
4. Hartley JL, Temple GF, and Brasch MA. DNA cloning using in vitro site-specific recombination. Genome Res 2000 Nov; 10(11):1788-95
5. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 2009 April; 6:343-345
6. “PDMS Process Notes” (June 30, 2008), Stanford Biomimetics and Dexterous Manipulation Laboratory, Available: http://bdml.stanford.edu/twiki/bin/view/Rise/PDMSProceSS
7. Saharinen, P., Vihinen, M., and Silvennoinen O. Autoinhibition of Jak2 Tyrosine Kinase Is Dependent on Specific Regions in Its Pseudokinase Domain. Mol Biol Cell., 2003 April, 14(4):1448-1459

Technology Transfer

Foreword.

The overarching aim of the project has been to develop a biosensor array based on the principles of mammalian olfaction. Using the tools of synthetic biology, we worked to create living cells that would serve as sensor elements in such an array and that would possess a fast, phosphorylation based memory circuit, responsive to G protein-coupled receptors (GPCRs) and histidine kinases (HKs) inputs (Fig.1). The phosphorylation based signal response allows a very fast biological readout, so that the biosensor can function in real time, unlike those based on transcriptional readouts. The toggle switch design incorporated into the circuit allows cells to maintain a memory of analyte exposure, which enhances the sensitivity of the sensor to analyte concentration. The use of GPCRs as the analyte receptors in the sensor elements allows enormous versatility in our ability to tune the array to any of a myriad different analytes. Moreover, by using different receptors with overlapping analyte specificities, we can detect a significantly greater number of analytes than the number of distinct sensor elements and can distinguish and quantify individual components presented in complex mixtures. Thus, the format matches the needs presented by real world conditions.

A second facet of the program has been a computational study to develop software that would decode the output of the biosensor. As in the olfactory system, each of the elements of the biosensor array responds to a distinct set of multiple analytes that overlaps the sets recognized by other elements. Accordingly, decoding what analytes were present in an applied mixture that produced a particular pattern of element responses becomes substantially non-trivial as the number of analytes in a mixture increases. Accordingly, a biosensor array is useful only if the pattern observed can be interpreted to yield the identities of the components in the mixture applied to the array. The Bayesian computational method we developed provides identification and quantification of all the components of a mixture tested in an array format and can be generalized to any olfactory-like biosensor array. This computational package also guides the design of any array to optimize the discriminatory capability and to minimize the array components.

As described in this final report, we made excellent progress on the first aspect of the project and completed the second task. Moreover, we have substantially streamlined the synthetic biology process by expanding the toolkit with which we create de novo designed strains. These results will facilitate further applications of synthetic biology to creating specific biology based circuits and sensors and have brought us to the point of reducing the biosensor to practice.

List of Figures and Tables.

Table 1. G-protein coupled receptors functionally expressed in yeast.

Problem Addressed

This project focused on development of a format for cell based biosensors, addressing three critical current shortcomings: 1) engineering the back end of the biosensor by creating a rapid readout of sensor activation; 2) engineering the front end of the biosensor by identification and implementation of suitable receptor elements that would provide broad spectrum coverage of the chemical space of interest; and 3) designing the sensor "brain" that would interpret sensor output to reveal the identities of the sensor inputs and quantify their amounts.

The first part of this program addressed a major problem in designing cell based biosensors, namely, to design a cell based system with a rapid readout. All previously described cell based reporter systems have used a transcriptional readout, which provides colorimetric, fluorimetric or growth readouts. These have proved useful in engineering and optimizing various biological circuits using the tools of synthetic biology and as a format for cell based assays for various drug screening purposes in the pharmaceutical industry. However, transcriptional based readouts are inherently slow, due to the multiple biological steps required for producing the final reporter product. Thus, in order to create a cell based assay that would provide useful feedback in real time, one needs a new platform for cell response that would transmit information on the presence of a stimulus and provide a detectable output in a very short time. To solve this problem, we proposed to develop a novel signaling and response network in the yeast *Saccharomyces cerevisiae* based on protein phosphorylation. The rapid in vivo kinetics of protein phosphorylation and the extensive information on natural phosphorylation networks in cellular signaling suggested that this was a feasible approach to solving this problem.

The second problem we addressed in this project was design of the front end of the sensor to allow broad spectrum coverage of chemical space. We proposed to approach this issue by basing the sensors on the family of G-protein coupled receptors (GPCRs). Currently, more than 4000 different GPCR genes have been identified, with specificity for an equally broad number of different chemical compounds. This diversity often allows selection of an individual receptor from the existing repertoire to fit a particular detection need. Moreover, we have recently shown that this diversity can be artificially increased by application of the tools of protein engineering to evolve a particular receptor to recognize a new chemical entity. Finally, the olfactory class of GPCRs exhibit degenerate and overlapping ligand recognition. This degeneracy is the basis of mammalian olfactory perception, allowing a relatively small number of distinct receptors (200-500) through a combinatorial process to recognize and distinguish a very large number (>100,000) of distinct chemical entities. Accordingly, we proposed to design the front end of our biosensor on the basis of the olfactory principle to allow maximum flexibility in application of the biosensor. This aim thus required that we engineer our yeast cell based system to functionally express a broad spectrum of GPCRs and to couple those receptors to the phosphorylation-based signaling network described above.

Our proposed use of combinatorial sensor arrays to detect a large number of analytes using a relatively small number of receptors raised a final problem that had to be addressed, namely, how to interpret the output of such a sensor array to identify the impinging chemical entities. The complex pattern of receptor responses to even a single analyte, coupled with the nonlinearity of responses to mixtures of analytes, makes quantitative prediction of compound concentrations in a mixture a challenging task. While the output of these cross-specific arrays in response to single compounds can generally be

interpreted through pattern recognition algorithms, computational analysis becomes more difficult when the array is presented with a mixture of compounds. Indeed, the non-linear nature of sensor responses to multiple ligands makes it hard to train discriminatory algorithms on a ‘‘typical’’ subset of patterns. The non-linear dependence of sensor output on ligand concentrations is generic in reporter systems and may be compounded by potential binding interference of the two ligands, saturation of the sensor output and, of particular concern, potential antagonistic action of one ligand on another’s activity. As a result, responses to complex mixtures have primarily been used to ‘‘fingerprint’’ specific mixtures rather than identify their constituents quantitatively. Accordingly, to achieve success, we needed to develop a novel and robust method for interpreting the output of the sensor arrays.

Major achievements

Enhancement of the tools of synthetic biology: rapid implementation of “plug and play” modules.

Achieving our goal of developing a cell based sensor with rapid response time required extensive application of the tools of synthetic biology, that is, of creating novel combinations of genes in a living cell that would redirect the normal function of the cell to perform a novel task. Given the large number of manipulations often required to re-engineer a cell to a desired novel specification, we spent some effort in developing methods to facilitate such manipulations.

To ease the time consuming process of large circuit construction in *Saccharomyces cerevisiae*, we have designed, built, and finalized a DNA assembly system for yeast systems. Our approach harnesses the strengths of yeast homologous recombination, a strategy employed for decades in biological research, and couples it to recent advances in synthetic biology stemming from recombination-based cloning strategies [4] and Gibson DNA assembly [5].

The system operates in two stages and pictorially represented in Figure 2:

1. Establishing a single transcriptional unit of promoter and gene.
2. Assembling multiple transcriptional units together into a backbone

For stage one, we start with a standard library of promoters and genes flanked by Invitrogen Gateway attL and attR sites. This library is fully compatible with other parts libraries, notably work of Susan Lindquist’s group at the Whitehead Institute and the plasmid libraries maintained by Harvard Medical School and Arizona State University. The Gateway ‘LR Reaction’ is performed in order to assemble promoter and gene pairs together into a destination plasmid such that the promoter-gene pair is flanked by defined 40-bp sequences. For example two ‘LR Reactions’ might yield the following plasmids: (Seq1 – Promoter – Gene – Seq2) and (Seq2 – Promoter – Gene – Seq3).

For stage two, we utilize the recently published Gibson Assembly relying on sequence and ligation independent cloning. By having 40bp homology regions on ends of individual DNA fragments, the Gibson reaction allows these 40bp flanked fragments to be combined to yield a single part. Thus we linearize the corresponding plasmids of stage one (Seq1 – Promoter – Gene – Seq2) and (Seq2 – Promoter – Gene – Seq3) and add it to a reaction containing a linearized vector (Seq1 – Vector – Seq3). Upon reaction completion, we obtain a circular vector containing (Seq1 – Promoter – Gene – Seq2 – Promoter – Gene – Seq3). A pictorial representation of this process appears in Fig 2B.

Under the aegis of this program, we have built the infrastructure necessary to implement this strategy, including a family of 18 promoters, 25 genes, and 8 gibson-compatible backbones. The promoter

library spans both constitutive and inducible promoters and the gene library includes fluorescent proteins, exogenous yeast components used in previously published *S. cerevisiae* papers from the Weiss laboratories. The Gibson-compatible backbones include the centromeric pRS plasmids allowing low copy propagation, the 2 micron plasmids allowing high copy propagation, and a novel site-specific integration plasmid. Figure 3 shows the construction of a family of 4 plasmids each containing a different fluorescent protein.

Creation of a rapid readout for cell based receptor activation.

A major goal of this project was to develop a means of rapidly detecting receptor activation in yeast. We have accomplished that goal by adapting the rapid phosphorylation cascade underlying MAP kinase signaling in yeast to yield a fluorometric response to receptor activation. As shown in Figure 4, three endogenous mitogen-activated protein kinase (MAPK) pathways – the mating pathway, the high osmolarity response pathway (HOG), and the filamentation pathway - coexist and function independently in yeast cells. In particular, the phosphorylation cascade of the HOG pathway results in rapid translocation of the Hog1 transcription factor from the nucleus to the cytoplasm in response to activation of the osmo-responsive receptor Sho1. While one can follow the translocation of Hog1 from the cytoplasm to the nucleus in microfluidic devices (Figure 5), this detection does not lend itself to use as a readout in a multi-element sensor array. Accordingly, we have designed and implemented a split GFP format so that activation of the receptor converts cells from non-fluorescent to fluorescent upon pathway activation.

Upon activation of the HOG pathway, the Hog1 transcriptional activator translocates into the nucleus, where it physically associates with the chromatin bound Hot1 protein. Accordingly, we created a strain in which Hog1 is fused to the N-terminal domain of the fluorescent protein eGFP and Hot1 is fused to the C-terminal domain of eGFP. The logic of this design is that in the absence of stimulation, Hot1 and Hog1 reside in different cellular compartments and, as a consequence, the two halves of GFP cannot associate and the cells are non-fluorescent. Upon stimulation, Hog1 relocates to the nucleus, binds to Hot1, allowing the two halves of GFP to associate and fold into a fluorescent protein, rendering the cells fluorescent. In this manner, pathway activation can be detected and quantified in whole cells or cultures of cells by the acquisition of fluorescence in proportion to the degree of stimulation. As shown in Figure 6, we have been able to accomplish this goal. Within five minutes of stimulation, we observe significant fluorescence of cells. This is more than an order of magnitude more rapid than any transcription based reporter assay described to date. In obtaining this rapid cell response, we have achieved one of the major goals of this project.

Coupling GPCR activation to our rapid response readout.

To functionally couple different GPCRs to our rapid readout, we have exploited the yeast cell's endogenous GPCR signaling pathway. Haploid yeast cells express a single GPCR, which is activated by pheromones produced by cells of the opposite mating type and which upon stimulation activates a MAPK pathway resulting in transcriptional activation through the Fus1 transcriptional activation (Figure 4). By engineering the G-protein that bridges the GPCR with the MAPK signal cascade, we have been able to create yeast strains that can functionally couple a variety of different mammalian GPCRs to the MAPK and thereby generate yeast cells that can detect and report on activation of mammalian receptors by their cognate ligands. This list of mammalian GPCRs that have been successfully coupled to the MAPK activation in yeast is provided in Table 1.

While our previous work provides a means of using a variety of GPCRs as the front end of a biosensor,

the readout for activation of these receptors has been transcriptional activation. To convert such strains to a rapid readout response, we have exploited our prior success in redirecting the pheromone signaling response into HOG pathway activation. As noted in Figure 4, the pheromone MAPK pathway shares a number of components with the HOG pathway. We previously showed that these two pathways were insulated from each other by mutual inhibition. Accordingly, by eliminating the mutual inhibition, we have been able to channel activation of the GPCR pathway into a HOG pathway response. Since this cross inhibition is mediate by the terminal MAP kinases of the pheromone pathway, Fus3 and Kss1, we have created variants of our reporter strains that lack the genes for both these kinases. In such strains, output of all three MAPK pathways is redirected solely to the high osmolarity pathway (Fig.2). Both, stress response and GPCR activation result in Hog1 phosphorylation and its translocation to the nucleus. This aspect of the program is still undergoing optimization.

Engineering a phosphorylation based toggle switch in yeast.

As a refinement of our cell based detector assay, we have focused on implementation of a toggle switch in the response pathway. By incorporating a toggle switch into our array design, we can generate a biosensor that possesses intrinsic memory. That is, the readout of the cell once exposed to an activating ligand will remain on even after the ligand is removed. This feature allows both assessment of prior exposure to a chemical entity as well as a means of resetting the detector at will. This toggling is achieved by negative coupling of two response pathways, each possessing a positive feedback loop. Accordingly, once one pathway is activated, it both remains activated and inhibits the activity of the second pathway. Stimulation of the second pathway turns off the first and, once this second pathway is activated, it remains activated.

We have proposed implementing a toggle switch by negatively coupling the pheromone responsive MAPK pathway to a Hog1p-responsive JAK2-STAT5 signaling module capable of feeding back to the phosphorelay input of the HOG1 pathway (Figure 1). Our toggle switch involves Jak2 inactivation with phosphorylation of Hog1 and STAT5 activation with Ypd1 phosphorylation. Central to this design are two chimeric proteins: Jak2-Hot1 and Stat5-HKRR. Active Hog1 fused to a phosphatase domain of SHP1, PTP, contacts and inactivates Jak2 as a part of the split-GFP reporter system. When Hog1 is not phosphorylated, Jak2 remains active and signals to Stat5. To enable phosphorylated JAK2-STAT5 to control Ypd1/Ssk1 phosphorylation we designed a Stat5-HKRR fusion protein where cytoplasmic Sln1 is fused to the C-terminal of Stat5. Sln1 is a 1220 amino acid protein with four distinct regions: (*a*) an N-terminal extra-cellular domain (ECD), (*b*) a cytoplasmic linker region, (*c*) a histidine kinase (HK), and (*d*) an aspartate response regulator (RR) domain [3]. We assume that activation of Jak2 dimerizes two Stat5-HKRR histidine kinase domains. Therefore, activation of JAK2-STAT5 will enable autophosphorylation of two HKs, resulting in suppression of HOG1 pathway through phosphorylated Ypd1 and Ssk1.

To achieve the toggle network topology, the JAK2-STAT5 signaling module must be made responsive to Hog1p. This link is made by producing the chimeric proteins PTP2-Hog1, Jak2-Hot1 and Stat5-HKRR. Sensitivity analysis of a quantitative model of the toggle network has shown that the activation of Ypd1 via Jak2 and Stat5 is critical to bistability. Our initial results showed that Jak2/Stat5 was capable of a 2-fold steady state increase of phosphorylated YPD1 mediated transcription. However, if this steady state fold difference can be increased, the regime of bistability increases dramatically.

The Jak2 protein is comprised of two domains, JH1 and JH2. JH1 is the catalytic domain while JH2 functions as a regulatory domain, inhibiting JH1's function [7]. Thus, we have created a JH1/Stat5 system where instead of utilizing Jak2, JH1 is substituted. Both Jak1/JH1 and Stat5 are under inducible

control. Figure 7 shows steady state population level mean fluorescence values. By utilizing JH1, we are able to get a 5.3 fold difference in the steady-state pathway activation. Further experiments are underway to modulate the JH1 and Stat5-HKRR levels to increase apparent gain for use in the toggle network.

Computational Methods for Array Data Deconstruction.

We have addressed a significant and unresolved problem in the development of array based detectors. Many array-based detector formats have been described but all are based on assembling a collection of individual sensor elements, each with non-identical but overlapping sensitivities to the collection of compounds or analytes that the experimenter wishes to detect and quantify. The fundamental problem is how to extract from the pattern of responses of the individual sensors on the array the nature of the stimulant applied to that array. The standard approach to this problem has been to invoke the ability to recognize patterns of activation. That is, if a test compound elicits a particular pattern of activation of a subset of sensors then one concludes that the test compound is equivalent to a reference compound that elicited the same pattern of activation. While this might be adequate for determining single test solutions containing a single compound, this approach is completely inadequate for determining what is present in a mixture of two or more compounds. This is even more problematic if the activities of the compounds on a sensor are not strictly additive, as would be the case for receptor antagonists. As far as we can determine, no one has developed a method for extracting from array data the composition of mixtures of compounds. This is unfortunate, since most of the real-life situations in which such arrays would be used would involve mixtures of compounds.

We have developed a Bayesian-based computational method for extracting the identities and amounts of compounds in a mixture using array based detectors. We previously used directed evolution to generate a collection of variants of the human UDP-glucose receptor, each of which possesses over-lapping but not identical binding and response activities against four separate ligands, UDP-glucose, UDP-galactose, UDP and UDP-glucosamine [5]. Using the yeast-based, transcriptional reporter for measuring GPCR activation, we first calibrated each of the receptors by determining dose response curves for each receptor against each of the ligands. We then applied nested sampling and Bayesian inference to the data to extract a binding affinity and an efficacy value for each of the ligands on each of the receptors. These receptors along with the calibration data constitute the detector array that we used for determining the nature of unknown mixtures.

In the second stage, we prepared various mixtures of the four ligands and applied them to the array. To determine the composition of a mixture, known dilutions of the sample were applied to each of the receptors and the resulting activation determined across the dilution series. These data are then processed by our Bayesian algorithm though nested sampling, which returns the most likely combinations and concentrations of ligands that would yield the observed response curves. This approach was remarkably effective in identifying the compounds present and their concentrations in a variety of different mixtures. We were accurate to within 20% in determining the presence and concentrations of any mixture of the four ligands. This is a spectacular and unprecedented result that will be broadly applicable to any array based detector system.

References

1. McClean, M. N., Mody, A., Broach, J., and Ramanathan, S., Decision Making in MAP Kinase pathways. *Nat. Genet.*, 2007, *39*, 409-414.
2. Magliery, T. J. et al., Detecting Protein−Protein Interactions with a Green Fluorescent Protein

Fragment Reassembly Trap: Scope and Mechanism. *J. Am. Chem. Soc.*, 2005, 127 (1), 146–157.
3. Hohmann, S., Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev*, 2002, 66(2):300–72.
4. Hartley JL, Temple GF, and Brasch MA. DNA cloning using in vitro site-specific recombination. *Genome Res* 2000 Nov; 10(**11**):1788-95
5. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 2009 April; 6:343-345
6. "PDMS Process Notes" (June 30, 2008), *Stanford Biomimetics and Dexterous Manipulation Laboratory,* Available: http://bdml.stanford.edu/twiki/bin/view/Rise/PDMSProceSS
7. Saharinen, P., Vihinen, M., and Silvennoinen O. Autoinhibition of Jak2 Tyrosine Kinase Is Dependent on Specific Regions in Its Pseudokinase Domain. *Mol Biol Cell.,* 2003 April, 14(4):1448-1459

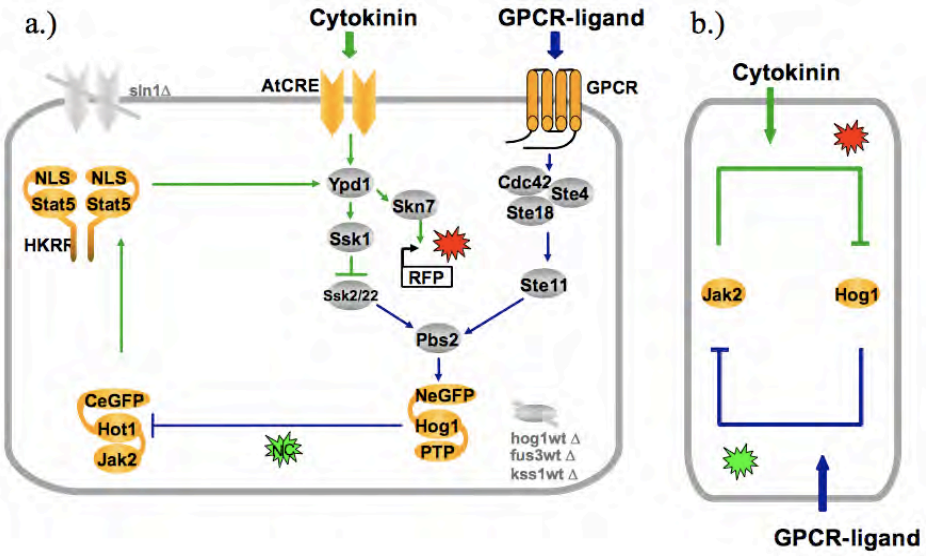

Figure 1. Toggle switch design. (a) detailed and (b) simplified representation; grey denotes endogenous and orange - exogenous proteins.

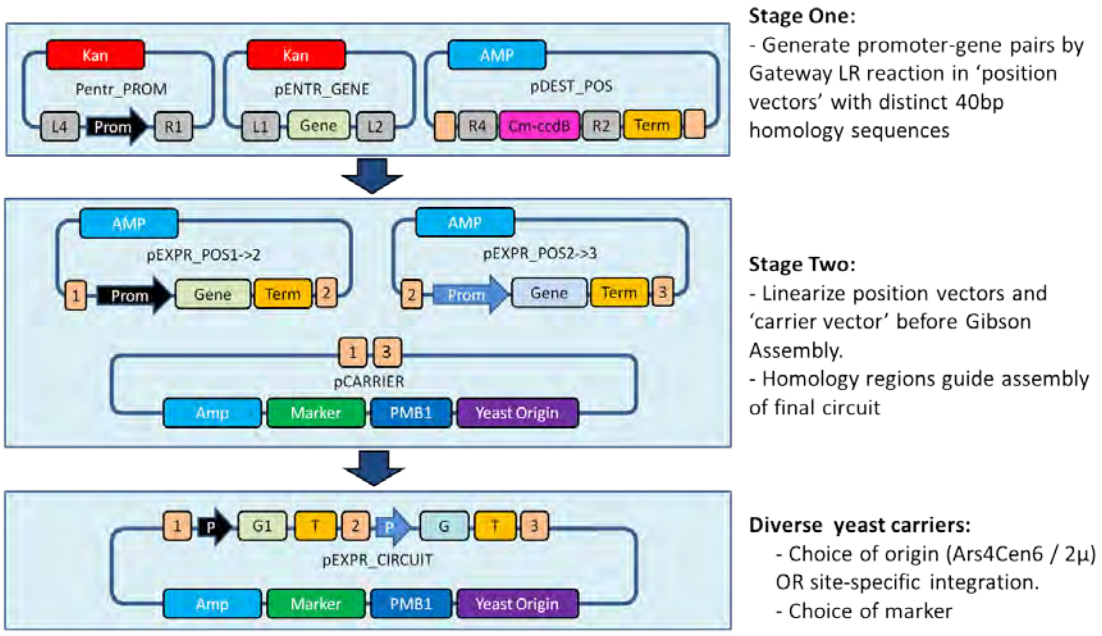

Figure 2 – Overview of large circuit construction. In stage one, Gateway recombination reactions are performed to assemble promoter:gene pairs into position vectors. Promoters and genes are in a standard entry vector format, compatible with internal and external libraries. In state two, the transcriptional units are linearized and combined with a chosen 'carrier vector' to yield a fully assembled circuit. Promoters or genes can be swapped via new LR reactions and circuit characteristics (selection or copy number / integration) can be chosen via new Gibson reactions with those carrier vectors.

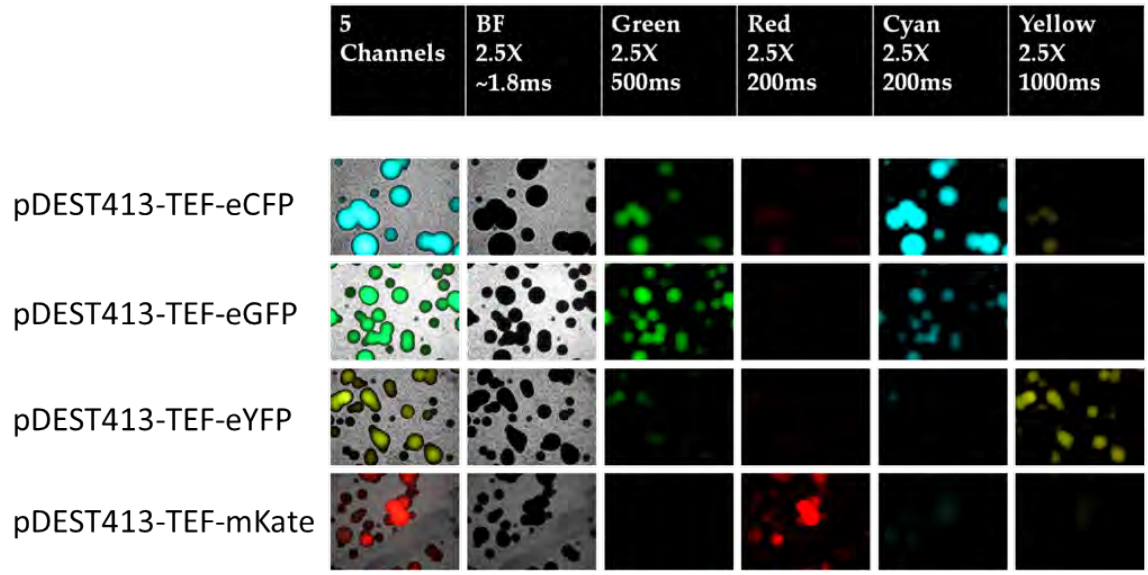

Figure 3 – Rapid circuit construction in yeast. Demonstration of stage one capabilities showing family of LR plasmids with various promoter and gene transcriptional units. Circuit design, construction, and introduction into yeast proceeded over 96 hours.

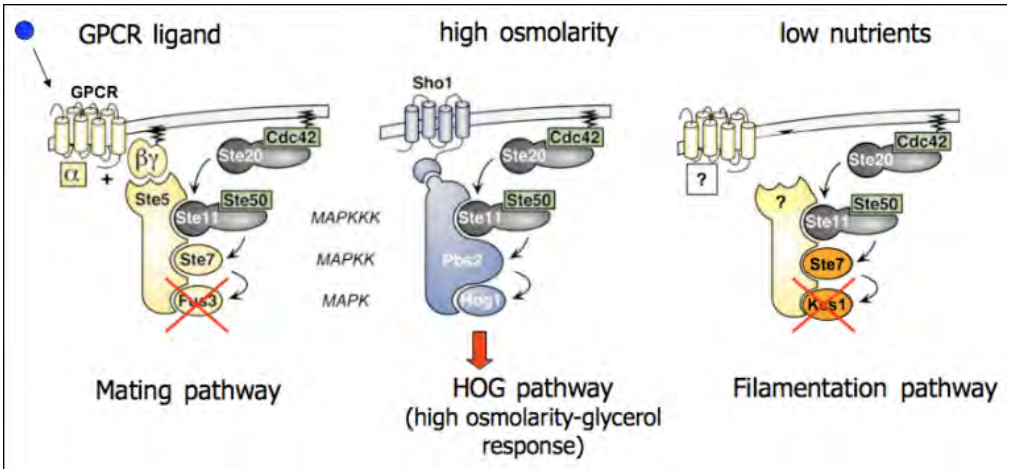

Figure 4. MAPK pathways in yeast. Mating factor receptor, Ste2, is replaced with human GPCR and Fus3 and Kss1 are deleted in order to redirect GPCR signaling exclusively to Hog1.

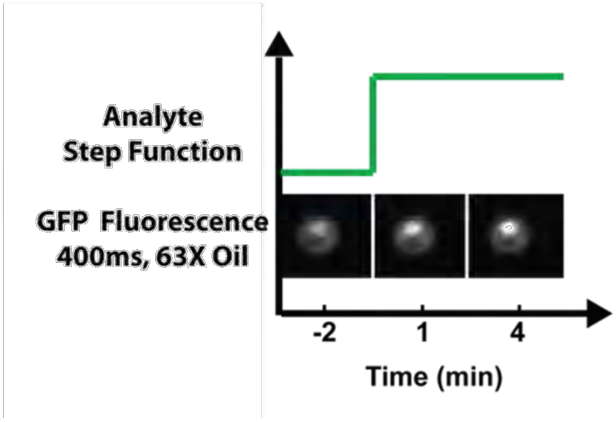

Figure 5 – Verification of Hog1-eGFP phenotype and single-cell imaging in microfluidic environment. Yeast strain YTS2ab_1 has constitutive Hog1-eGFP production and thus upon a step function of sorbitol, we observe nuclear localization of the fusion construct. YTS2ab_1 - W303-A background, *hot1D::loxP, hog1D::loxP*, *HO::Hog1:Hog1-eGFP*

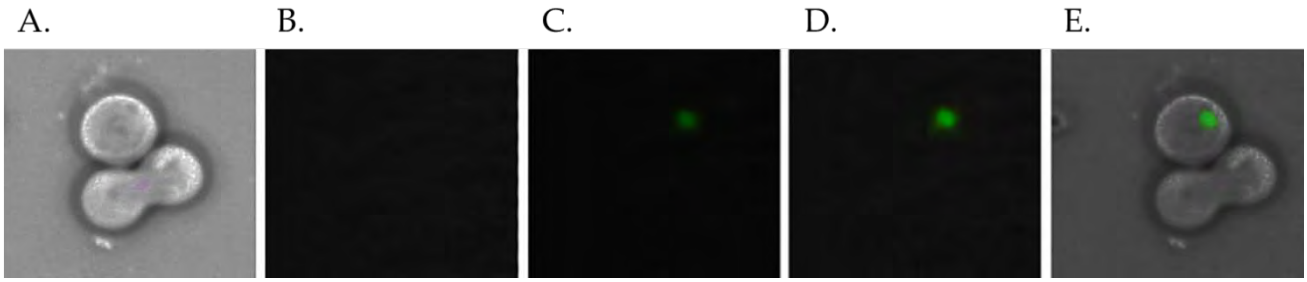

Figure 6 – Verification of Hog1-NeGFP and Hot1-CeGFP interaction and reconstitution.
Yeast strain YTS2ab_3 has constitutive Hog1-NeGFP and Hot1-CeGFP. We expect a sorbitol pulse to cause Hog1-NeGFP to localize to the nucleus, and the resulting Hog1-Hot1 interaction to drive nuclear fluorescence. YTS2ab_3 – W303-A background, *hot1D::loxP, hog1D::loxP, HO::Hog1:Hog1-NeGFP_Hot1:Hot1-CeGFP*
Time = 5 min prior to Sorbitol Pulse (A) Brightfield, 63X Oil (B) GFP Channel 400ms exposure
Time = 5 min post Sorbitol Pulse (C) GFP Channel
Time = 15 min post Sorbitol Pulse (D) GFP Channel (E) Brightfield, GFP Overlay

Plot Label	Genetic Circuit			
Blank	TM182α			
JAK2	TEF:rtTA	TEF:kanMX	GALS:mJak2	TR-SSRE:eGFP
JH1	TEF:rtTA	TEF:kanMX	GALS:JH1	TR-SSRE:eGFP
STAT5-HKRR	TEF:rtTA	tetO7CYC1min:mSTAT5-HKRR	TEF:kanMXC	TR-SSRE:eGFP
JAK2 & STAT5-HKRR	TEF:rtTA	tetO7CYC1min:mSTAT5-HKRR	GALS:mJak2	TR-SSRE:eGFP
JH1 & STAT5-HKRR	TEF:rtTA	tetO7CYC1min:mSTAT5-HKRR	GALS:JH1	TR-SSRE:eGFP

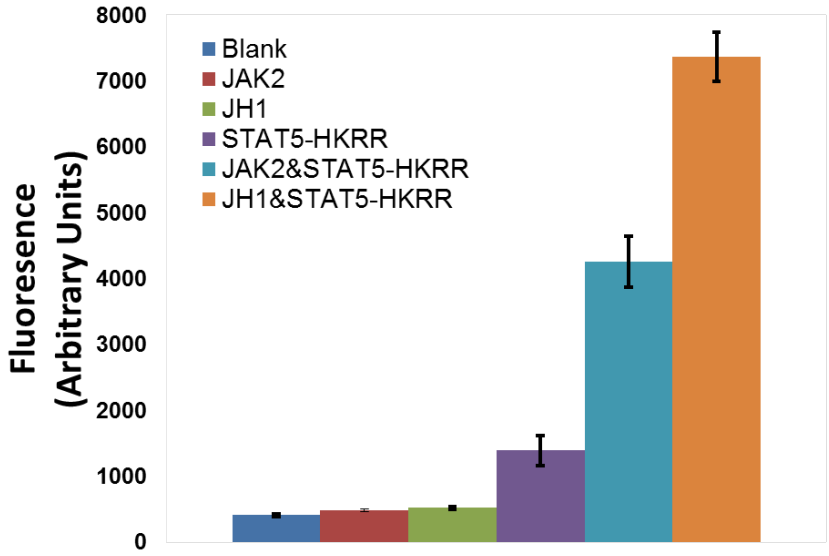

Figure 7 – Indirect transcriptional readout of Jak2/JH1 & Stat5-HKRR mediated-activation of YPD1. The strains tested are all in TM182alpha background (sln1::hisG, pSSP25) and contain HO integrated circuits as follows in liquid media supplemented with 1.0ng/mL DOX and 2% Galactose

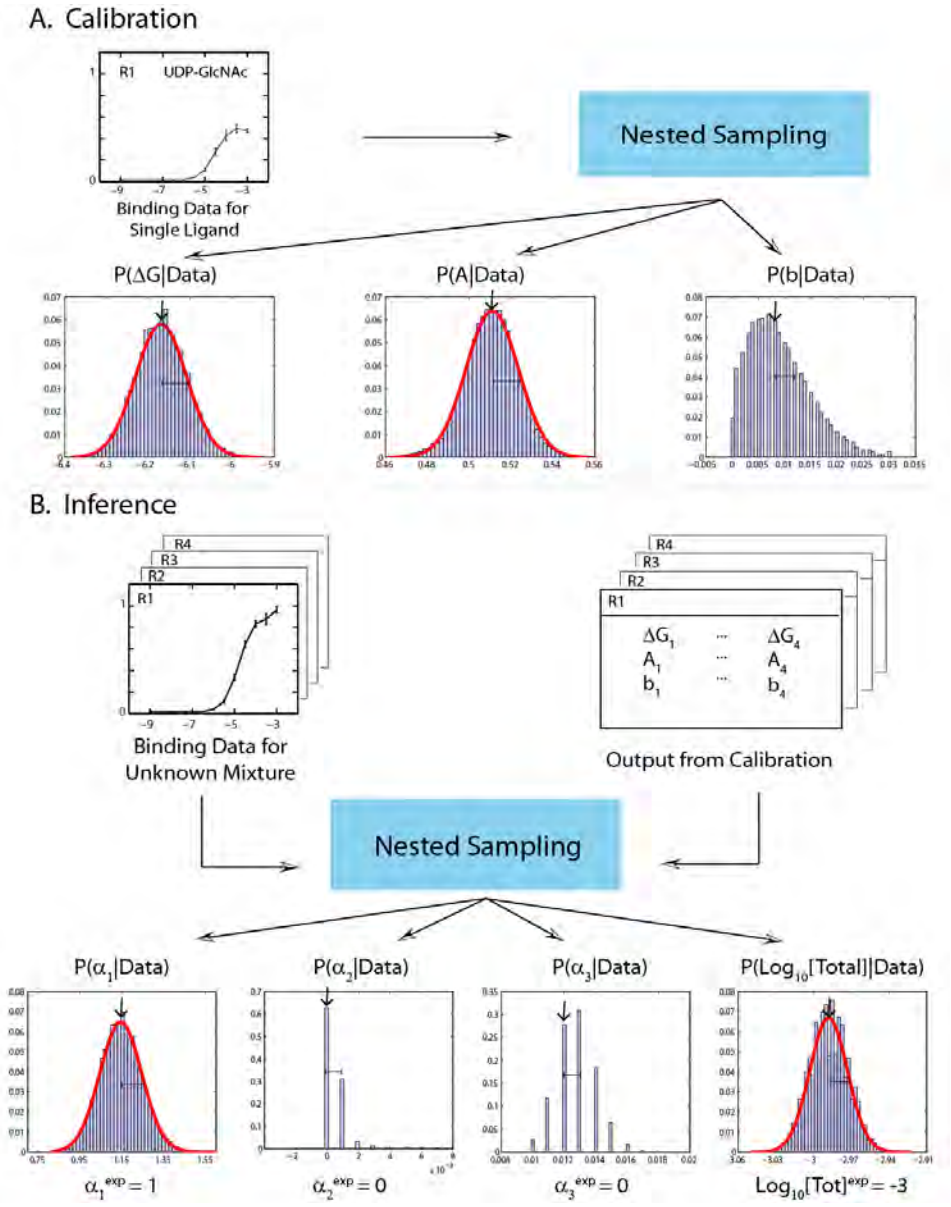

Figure 8. Bayesian inference method to determine complex mixtures from array data. A. Calibration. B. Inference

Table 1. G-Protein Coupled Receptors Functionally Expressed in Yeast

Vasointestinal peptide-1 (VIP-1)
Nociceptin 2
Bombesin-3 (BRS-3)
Adenosine A1
Adenosine A2a
Somatostatin-1
Somatostatin-2
Somatostatin-3
Melanocortin-4
Neurotensin-1
Neurotensin-2
CRF2α
CRF1
Calcitonin gene related peptide
adrenomedullin
Vasopressin-2
GRP
Orexin-2
EDG-1 (Sphinogsine-1-P)
KIAA0001 (UDP-glucose)
MasSB (substance P)
C5a
IL8 (CXCR1, CXCR2)
Thrombin
Melatonin-1a
Melatonin-1b
Melatonin-1a like
FPRL1
FPR-1
NPY-Y1
NPY-Y2
HGalR1
CCR5
Edg-2 (lysophosphotidic acid)
Muscarinic M1
Muscarinic M3
Muscarinic M5
Somatostatin-4
Somatostatin-5
VPAC1
Edg-3, 4, 5, 6, 7